DOCUMENT RESUME

ED 431 218 EA 029 863

AUTHOR Grip, Richard S.; Young, John W.
TITLE Predicting School Enrollments Using the Modified Regression Technique.
PUB DATE 1999-04-00
NOTE 28p.; Paper presented at the Annual Meeting of the American Educational Research Association (Montreal, Quebec, Canada, April 19-23, 1999).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Comparative Analysis; Elementary Secondary Education; Enrollment Management; *Enrollment Projections; Models; *Regression (Statistics); *School Demography; School Statistics; Statistical Analysis
IDENTIFIERS New Jersey

ABSTRACT

This report is based on a study in which a regression model was constructed to increase accuracy in enrollment predictions. A model, known as the Modified Regression Technique (MRT), was used to examine K-12 enrollment over the past 20 years in 2 New Jersey school districts of similar size and ethnicity. To test the model's accuracy, MRT was compared to the Cohort-Survival Ratio method--the most common method employed by school planners--using two prediction time frames, 1990-96 and 1993-96. These time frames were then compared to actual enrollments. The results show that MRT performed significantly better than the Cohort-Survival Ratio, achieving much lower absolute-mean-error rates. Error rates were larger for one of the districts using either method; for most accurate results, both models worked best under conditions of uniform growth or decline. Furthermore, smaller data sets produced the best results, indicating that, rather than picking the greatest number of years available, school planners should use data that closely resembles prevailing trends. It is suggested that MRT can provide school planners with a more accurate methodology for predicting long-term school enrollments, helping them to make decisions regarding the opening and closing of schools, hiring new faculty, and planning time. (RJM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 431 218

Predicting School Enrollments
Using the Modified Regression Technique:

Richard S. Grip

and

John W. Young

Rutgers University

Running Head: The Modified Regression Technique

Contact Information:
Telephone: (732) 297-4506
E-mail: rsg@statforecast.com

This paper was presented at the annual AERA conference, Montreal, Canada, April 1999

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

R. Grip

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

EA 029 663

BEST COPY AVAILABLE

Abstract

Predicting public school enrollments continues to be a challenge to school planners due to the number of different methods available for making predictions and the numerous variables that influence school enrollments. The most common method employed by school planners, the Cohort-Survival Ratio method, is effective in predicting enrollments for 1 to 3 years into the future. However, its effectiveness as a long-term predictive tool is poor. A new model, the Modified Regression Technique was used to effectively predict enrollments. The Modified Regression Technique utilized three variables, the number of students in a grade, migration, and housing growth, which have a significant impact on school enrollments.

The K-12 enrollments from the past 20 years of two school districts of similar size and ethnicity were used in this study. To test the accuracy of the methods, enrollments were calculated for two prediction time frames, 1990 to 1996 and 1993 to 1996, and were checked against actual enrollments. Absolute Mean Error Rates were calculated for each school district for the prediction time period. The results showed that the Modified Regression Technique performed significantly better than the Cohort-Survival Ratio by achieving much lower Absolute Mean Error Rates. It is proposed that the Modified Regression Technique can provide school planners with a more accurate methodology for predicting long term school enrollments. This will assist school administrators in making accurate decisions regarding opening and closing schools, hiring new faculty, and will allow sufficient time to plan accordingly to achieve these goals.

The Modified Regression Technique
A New Method for Public School Enrollment Projections

Until a decade ago, school facility planners had sufficient time to prepare for changes in enrollment in the public school systems. However, in recent years, changes have occurred so quickly in many of America's schools that it has been difficult to make the necessary adjustments to accommodate the rapidly changing student population. Knowledge of the changes occurring within the student populations is imperative for facility planning, staffing, determining participation rates for curriculum programs, and for opening or closing of schools. Administrators require sufficient lead time to be able to prepare for the changes that lie ahead. An accurate prediction model of school enrollments is necessary in order to exercise appropriate planning strategies for the future. Currently, the most common prediction model, known as the Cohort-Survival Ratio (CSR), utilizes a "calculation of a series of survival rates that indicate the fraction of students in one grade in a given year who progress to the next grade in the next year" (Shaw, 1984, pg. 53). Although this is very accurate in short-term planning, its accuracy as a long-term predictive tool is poor to questionable. The biggest advantage of using the Cohort-Survival Ratio method is the simplicity in calculating enrollments and the low cost of obtaining the actual enrollments needed to calculate the progression ratios.

Another technique which is not currently used by school planners but is easy to implement for enrollment projections is the moving average. A moving average computes an average of the most recent data in a time series and uses the average to forecast future values. Moving averages have been utilized at the post-secondary level to calculate enrollments but there is little research showing the effectiveness at the secondary level (Pfitzner, 1987). A number of other models have been developed previously with varying degrees of success. These models include the Percentage-Survival Method of Forecasting, the Graphic Technique Method, Dwelling Unit Multiplier Technique, Markov Chain Models, Autoregressive Integrated

Moving Average (ARIMA), Population Ratio Technique, Multiple Factor Method, Regression Method, and Forecasting by Analysis.

The goal of this study was to construct a regression model to more accurately predict enrollments than calculated by either the Cohort-Survival Ratio method or by moving averages for a time frame of one to seven years into the future. While a regression model may be more difficult to construct than either the Cohort-Survival Ratio method or moving averages, it may be more accurate in predicting changes in enrollment.

Several studies (Shaw, 1984, Murdock & Ellis, 1991) have discounted the use of regression techniques in predicting enrollments. One reason researchers have not favored a regression approach is because separate regression equations are needed for grades K to 12 for accurate predictions and therefore each grade level's enrollment projection contains error which compounds when summing the total enrollment for all of the grades. Another reason is that many of the values for the independent variables are unknown for a future date of prediction and must be extrapolated from past values. Not only should the independent variables correlate highly with the dependent variable, but the variables should also exhibit robustness in extrapolation (Webster, 1971). Despite the apparent drawbacks of using a regression approach, other researchers using regression to predict enrollments have been quite successful. Chan (1993) stated that a regression approach is a "well balanced approach to enrollment forecasting"(p. 31). In a separate study to test whether regression analysis was a better method of prediction than the Cohort-Survival Ratio, Webster (1970) found a regression model to be significantly more accurate than the CSR. In that study, a simple linear regression equation was used with the number of students in a previous year being the sole predictor.

The investigator of this study hypothesized that a regression model, to be known as the Modified Regression Technique (MRT), with a series of predictors including the number of students progressing from grade i to grade $i + 1$, can greatly enhance the accuracy of enrollment projections. In enrollment predictions for the South Burlington, Vermont schools, Harris, Burrage, and Smith (1986) identified seven

key variables that were used in their projection. These included birth rates, death rates, migration into the community, construction of new homes and average number of occupants of each, and percentage of school-age children attending community schools.

Variables similar to those in the South Burlington study were employed in this study to develop the MRT and included the number of current students, birth rates, migration rates, and total number of housing units. Data for a twenty year period from 1973 to 1992 were used to predict enrollments for the years 1993 to 1996 for two suburban school districts in central New Jersey of similar demographic and socio-economic composition. A smaller subset of data from 1973 to 1989 was used to predict enrollments from 1990 to 1996. The larger data set was used to increase the sample size needed for an adequate regression model. The larger data set, however, may not reflect the most current enrollment trends. A smaller set using more recent data may better reflect the enrollment trends at the time of forecasting.

Variables

Birth Rates

In addition to the number of current students in a school district, the number of births in a given year is useful in predicting the number of kindergarten students five years into the future (Glass & Fulmer, 1991, Webster, 1971). Previously, birth rates have not been used in regression models since it is unknown what the number of births will be for a future time period. Assuming that the nation is not experiencing a tremendous baby boom as seen at the end of World War II, a simple linear or curvilinear regression model can be used to calculate the number of expected births. Caffarella (1987) stated that after a "baby boomette", which occurred in the United States from 1973 to 1988, there will be a declining number of births nationally up until the year 2000. When predicting the number of future births, other factors such as economic conditions, ethnicity changes, and housing growth or decline should be considered. Besides using quantitative measures to determine future birth rates, qualitative judgment should also be employed to modify the data.

Although birth data are readily available, a great drawback is the difficulty in effectively utilizing the data. Often, schools districts do not have the same geographical boundaries as the towns or municipalities they represent. For instance, a municipality may be divided where the east side of town is sent to X Regional School District while the west side of town is sent to Y Regional School District. From a planning perspective, it makes enrollment projections quite difficult since it may not be known (or may be known but too cumbersome to calculate) which new births will factor into the enrollment projections.

Migration

Perhaps the greatest source of variation in enrollment projections can be directly traced to students who migrate into or out of the district. According to Weldon, Hurwitz, and Menacker (1989), there is increasing mobility in American families where relocation occurs on average once every six years. While it is unknown what causes most families to move, the degree of transition in a community depends on whether the community is rural or urban, and on the racial composition.

A second factor which affects migration is economic conditions such as job creation or loss (MacConnell, 1957). In an industrial, manufacturing town, the closing of the town's factory may be devastating to a school system. If there is little work in surrounding communities or if parents working in this industry are skilled only in this field, residents may be forced to relocate to find employment. Alternatively, new industry in a town could provide a boost to school enrollment due to the number of workers and their families which would be added to the community. The variable used in the MRT to account for these changes is known as the migration ratio. The migration ratio, which is identical to the cohort survival ratio, is the ratio of the number of students in grade $i + 1$ to the number of students in grade i.

A final variable included in the MRT is the total number of housing units in the township being studied. One of the most important factors leading to inward migration into a school district is the presence of new housing developments (Glass & Fulmer, 1991). If a township has open land rapidly being converted into housing developments, this will inevitably lead to an increase in the number of students in

the school. The housing variable accounted for the total number of single-family or multi-family homes, apartment units, townhouses, and condominiums existing in a township. The number of new housing units was estimated through a linear regression equation since the number of future homes would not be known at the time the enrollment predictions were made. Qualitative judgment of the number of future homes was also employed since the amount of vacant land in each of the townships ultimately determines the number of new homes that can be built. Since the number of housing units was to be estimated, a high degree of error may result from using this variable.

To determine the accuracy of the regression models, the projections were compared with enrollment predictions for the same time period using the Cohort-Survival Ratio method and moving averages. Since the actual enrollments from the years 1990 to 1996 will be known, percent errors for each grade was calculated for each method employed. A comparison of the Absolute Mean Error Rates (AMER) from each of the three models was used to identify the model which best predicted future school enrollments.

Enrollment Projection Methods

The Moving Average

A simple time series technique employed in enrollment projections is the moving average. Moving averages are based on an average of the most recent data of a time series and then uses this value to forecast future measurements (Lee & Shim, 1990). The averages are plotted as a function of time and future values can be obtained by then extrapolating the trend visible on the graph. The moving average is calculated as follows:

$$\text{Moving Average} = \frac{\Sigma \text{ (most recent n measures)}}{n}$$

where n is the number of data measures.

Increasing the number of periods (n) that is used to calculate the moving average will smooth out random variation at the possible expense of accuracy in depicting the trend. Shorter number of periods used for calculations will help to accurately reflect the trend that is occurring (Lee & Shim, 1990).

Despite the ease in calculation and cost efficiency in collecting information for only one variable for the moving average, this technique has its disadvantages. First, moving averages do not provide reliable results when the data trend is increasing or decreasing very rapidly. Second, since this technique uses only a single variable, the number of students in each time period, the results do not provide an explanation as to why the enrollments are increasing, decreasing, or remaining steady (Pfitzner, 1987). Finally, the moving average can not predict cyclical changes. Instead, the moving average can only detect the general trend of the data.

Cohort Survival Ratio

In 1930, Dublin and Lodka provided an explicit age-breakdown which enabled analysts to follow each cohort through its life stages and apply appropriate birth and death rates for each generation (Ascher, 1978). A descendant of this process is the Cohort Survival Ratio Method (CSR) which is the most commonly used quantitative method to forecast public school enrollments. In this method a Cohort Survival Factor (CSF) is computed for each grade which essentially compares the number of students in a particular grade to the number of students in the previous grade during the previous year. The equation is usually expressed as:

$$\text{CSF} = \frac{E_{g,\,y} + E_{g,\,y-1} + E_{g,\,y-2} + E_{g,\,y-3} + \ldots + E_{g,\,y-n}}{(\text{from g-1 to g})\; E_{g-1,\,y-1} + E_{g-1,\,y-2} + E_{g-1,\,y-3} + E_{g-1,\,y-4} + E_{\ldots\ldots} + E_{g-1,\,y-n-1}}$$

where E is the actual enrollment for a specific grade (g) during a specific year (y). The CSF indicates whether the change between grades results in enrollment that is stable, increasing, or decreasing between grades. A CSF of one indicates a stable enrollment, less than one indicates declining enrollment, while greater than one indicates increasing enrollment (Caffarella, 1983). If, for example, a school district had 100 fourth graders and the next year only has 95 fifth graders, the CSF would be 0.95.

Although CSR is fairly simple to use, this technique makes several assumptions which can not be violated. First, the researcher assumes that what happens in the past will also happen in the future (Caffarella, 1983). Second, a cohort's net retention in the district or school from grade g-1 to g at time t will be exactly the same as for the next older cohort at a future time t + 1 (Bernhardt, Pullum, & Graham, 1983). In essence, this method provides a linear projection of the population (Castaldi, 1989). Unfortunately, many school facility planners use CSR without fully understanding when this technique is most and least appropriate. CSR is most applicable for districts that have relatively stable increasing or decreasing trends without any major unpredictable fluctuations from year to year. In school districts that are encountering rapid growth, CSR must be modified and supplemented with additional information to ensure accuracy (Cline, 1985). Dekel (1994) stated that CSR is appropriate for rural districts as these district are less susceptible to great changes in population due to migration. Also, CSR is only able to accurately forecast for a very short interval into the future. Researchers disagree on the range of effectiveness. Some researchers claim CSR has no predictive ability (Bernhardt, Pullum, & Graham, 1983), others claim one to two years (Glass & Fulmer, 1991), while others believe CSR is valid for less than five years (Colombo, Dekker, & Petronis, 1990) or as many as seven years (Shaw, 1984). Beyond one to two years, the projections tend to be highly suspect since CSR's major assumption is a linear trend which may not hold true after a few years (Chan, 1993, Glass & Fulmer, 1991). One outcome which can occur if the assumptions of using CSR are violated is that systemwide enrollment projections for the whole district may be relatively accurate, but the school by school enrollment within the district may be in error (Chan, 1993). In CSR's simplest form, no more than two years of past data should be used even if more is available (Bernhardt, Pullum, & Graham, 1983). Using three or four year trends is a more complicated approach that requires calculating an average progression ratio over the desired time period yet tends to smooth out any large variations.

The Modified Regression Technique

In this study, the independent variables were entered simultaneously in a regression model known as the Modified Regression Technique. Independent variables which correlated highly with each other were not included in the regression equation. A series of regression equations was formulated for each grade (K through 12) for the seven year period from 1990 to 1996 estimated from data from the years 1978 to 1989 and from 1973 to 1989. A second series of regression equations was formulated for each grade for the four year period from 1993 to 1996 based on enrollment data from the years 1973 to 1992 and from 1978 to 1992. From the literature, the Cohort-Survival Ratio method should provide adequate projections for a time period of one to three years into the future but accuracy beyond this time frame is unproven. However, this researcher theorized that the MRT would also provide accurate projections in the short range due to the use of additional important predictor variables. The MRT was used to predict enrollments from one to seven years into the future, a time span of prediction in which many other models faltered.

Implementation of the Modified Regression Technique

To implement the MRT, data from a twenty year period from 1973 to 1992 were used to predict enrollment for two separate time periods. First, data from 1973 to 1989 and 1978 to 1989 were used to predict enrollments for the years 1990 to 1996 which is a period of one to seven years into the future. Second, data from 1973 to 1992 and 1978 to 1992 were used to predict enrollments for the years 1993 to 1996 which is a period of one to four years into the future. The purpose of using two time periods to predict enrollments for the same time interval into the future is to use the data that best reflects the enrollment trends at the time of prediction. For instance, in predicting enrollments for 1990 to 1996 using data from 1973 to 1989 as opposed to 1978 to 1989, the data from the period 1973 to 1977 may include enrollment trends that are not consistent with more recent trends of enrollment at the time of projection. While school districts can be expected to have cyclical periods of increasing and decreasing enrollment, the researcher should use data similar to the most current trend in enrollment growth or decline occurring in the district. However, a drawback to using a smaller data set is that the validity of the regression equation

decreases. The purpose of using a larger data set (1973 to 1992) is to increase the sample size used to formulate a regression equation in order to obtain more accurate results.

While two predictions were made for a seven year period into the future, two other predictions were made for four years into the future. The researcher expects that the models predicting only four years into the future will be more accurate than the models predicting seven years into the future as the accuracy of the model will deteriorate as the forecast horizon is enlarged. Since the Cohort-Survival Ratio method becomes less accurate in predicting more than four or five years into the future, the researcher expects that use of the MRT for predictions of four years into the future will be as accurate as CSR but will surpass CSR's effectiveness when predicting seven years into the future. In a survey conducted by Schellenberg and Stephens (1987), only 2 of 50 districts surveyed made enrollment predictions for more than five years into the future. Most of these districts used CSR for enrollment projections indicating many school planners realize CSR is a poor long-term planning technique.

Enrollment predictions for the study end in 1996 as this was the most recent data available for comparison. Predicted enrollments were compared with actual enrollments for the same time period as a measure of accuracy. In effect, the study is a simulation of what would have been available to a researcher conducting the analysis in either 1989 or 1992.

As the enrollment prediction model uses data from the years 1973 to 1989 and 1978 to 1989 to forecast enrollments for the years 1990 to 1996, the number of new housing units for the future would also be unknown to a researcher in the year 1989. In the enrollment prediction models using data from 1973 to 1992 and 1978 to 1992 to forecast enrollments for the years 1993 to 1996, a researcher would not have known the number of new housing units for 1993. While a recession or economic boom can significantly alter the trend of housing growth or decline, a major assumption in determining future growth is that past housing growth will be a good predictor. A second assumption is that there is undeveloped land remaining in a township. If the amount of undeveloped land is expected to decrease dramatically in the future, growth of housing may not reflect historical trends.

To predict the number of new houses, condominiums, townhouses and apartment buildings built in 1990, for instance, regression equations were constructed based on housing data from the years 1973 to 1989 and 1978 to 1989. If the township had very little remaining undeveloped land in 1989, using a regression equation to predict the number of additional housing units to be built in 1990 is unrealistic. Qualitative analysis combined with quantitative analysis is essential in order to produce accurate results.

Another predictor variable was the number of students in a grade g in a given year. In the Cohort-Survival Ratio model, the best predictor of the number of students in a grade g+1 is the number of students in grade g. Webster (1970) used the number of students progressing from grade g to grade g+1 as the sole predictor variable in a simple regression equation. The Modified Regression Technique should give better results than Webster's model since it uses two additional independent variables. To predict the number of students at the kindergarten level for the years 1991 and 1992, actual birth data from the state health department of vital statistics from the years 1986 and 1987 were used. However, to predict kindergarten classes for the years 1995 to 1996, estimated birth rates were predicted from a separate regression equation based on the number of births from past years. Since part of the analysis was conducted ending with data from 1989, birth data for 1990 and beyond which determine kindergarten classes for 1995 and beyond, was treated as unknown. For the data which ends in 1992 and was used to predict enrollments for 1993 to 1996, no predicted birth rates were needed since births in 1992 would help determine the kindergarten class of 1997 which was not part of this study. Birth data was used to predict the number of kindergarten students only and was not used in calculating predicted enrollments for grades 1 to 12.

A third variable incorporated into the MRT is a migration ratio which is identical to the cohort survival factor. The formula to calculate the migration ratio is:

$$\text{Migration Ratio} = \frac{\text{(Students in grade } i+1 \text{ - Students in grade } i)}{\text{Students in grade } i} + 1$$

Migration ratios greater than one indicate that students are migrating into the school district while ratios less than one indicates students are leaving the district and are not progressing to the next grade. The

migration ratios were calculated per grade for the years of prediction. The researcher anticipated that the migration ratios will be greater than one in the elementary years as parents are more likely to change their children's school than in the high school years. Also, migration ratios less than one are expected for the upper grades as students entering high school from the middle school may transfer to parochial or private secondary school. For these reasons, a migration ratio was calculated per grade to account for the fluctuations that occur from grade to grade. For the predicted year of 1990, thirteen separate regression equations were computed, one for each grade. For 1991, the predicted number of children in grade i in 1990 was used in the regression equation as the current number of students in grade i + 1 for the year 1991. For instance, if the predicted number of students in the 4th grade in 1990 is 100, this figure is the number of students expected to be in the fifth grade in 1991. In reality, all 100 students are not expected to progress to the next grade. Some students may be retained, others may transfer out of the district while new students may enter the district. The migration ratio variable should account for much of the movement that occurs in a district on a yearly basis.

Methods

The grade level enrollments for two suburban school districts in central New Jersey were used in this study. Both districts are similar in size and racial/ethnic composition as shown in Table 1 below.

Insert Table 1 about here

In 1996, District A had a total of 5,686 students while District B had 6,241 students. As a test for accuracy, the enrollment projections from the regression models for both the four and seven year

projections were compared with the predicted enrollments for the same time period using the Cohort-Survival Ratio model and a three year moving average.

The Cohort-Survival Ratio model was selected as a comparative tool because this technique is the most popular method currently used by public school administrators in predicting short term and long-term enrollment. The computed progression ratios which were multiplied by the number of students in grade i to calculate students in grade i + 1 in the following year was calculated for a two year average and a four year average. In using a two year average, the most recent enrollment trends were utilized while the four year average was a conservative approach using more historical data in computing the average.

In calculating the enrollment from the moving average method, a three year average of actual enrollments was calculated and plotted on a graph as a function of time to study the existing trend. For instance, since the data used actual enrollments from 1973 to 1992, the first moving average averaged the enrollments for a particular grade for 1973, 1974, and 1975. The second moving average was for 1974, 1975, and 1976. In the first moving average model, this reiterative process continued for each grade through 1989 and the extrapolated values from the graph of enrollments versus time were used to predict enrollments for 1990 to 1996. In the second moving average model, the same steps were repeated except the averages were calculated up until 1992 so predictions could be made for the years 1993 to 1996.

Future enrollments were calculated using three different prediction methods: the Moving-Average Method, Cohort-Survival Ratio, and the Modified Regression Technique. The results using the moving average method provided very large errors not consistent with the results from using CSR or MRT and therefore will not be reported here. The complete enrollment projections using the moving average method is documented by Grip (1998). An additional short-term forecast for the four year period from 1993 to 1996 was predicted from enrollment data from the period 1978 to 1992. Also, an additional long-term forecast from 1990 to 1996 was calculated from data from the years 1978 to 1989. By using only a 10 or 15 year period of data to calculate enrollments, the researchers expected the data would more closely approximate the population trend at the time of projection. Using 15 or 20 years of data may not typify the

current population trends and can lead to less accurate results. One drawback of using fewer years is that a regression analysis is more reliable for sample sizes greater than 30. By calculating future enrollment for two time frames, 1990 to 1996 and 1993 to 1996, the researcher could detect whether a particular method was effective in the short-term or long-term and if the number of years used to predict enrollments affected the accuracy of the projections.

In the MRT, four multiple regression analyses were performed for each district. In each regression analysis, thirteen separate regression equations were calculated, one each for grades K to 12. For the analyses that predicted long-term enrollments from 1990 to 1996, birth data were predicted for the years 1990 and 1991 which would subsequently become the kindergarten classes of 1995 and 1996. Predicting births for the short-term predictions from 1993 to 1996 was not necessary since the births in 1992 would become the kindergarten class of 1997 which is not a year included in the study.

When calculating the future number of houses for either the time period 1990 to 1996 or from 1993 to 1996, District A results utilized simple linear regression equations while District B results used simple polynomial equations reflecting a faster rate of housing growth. The number of years of housing data used to develop the prediction was 12, 15, 17 or 20 years depending on which of the four regression analyses were being calculated. One difficulty encountered in predicting housing enrollment was that existing housing totals were only available for the years from 1978 to 1991. Housing data for each district from 1973 to 1977 had to be estimated by using the number of building permits issued in that particular year.

When using the MRT, the migration ratio variable was computed in the same manner as the Cohort-Survival Ratio. Migration ratios for each grade had to be predicted for a seven-year period for the long term-enrollments from 1990 to 1996 and for four years for the short-term enrollment regression equations for 1993 to 1996. When plotting the migration ratios for each grade as a function of time for each district, no distinct trend was evident. The migration ratios appeared to be a random function. Since a regression equation could not be used to predict migration ratios, an average of the migration ratios for

each grade for the last 5 years was calculated. This ratio was then used as the expected migration ratio in the years 1990 to 1996 or 1993 to 1996.

When developing the regression equations for each grade, all three independent variables, migration ratio, total housing units and current number of students, were entered simultaneously into the regression model. Often, the housing variable was only statistically significant for the early grades such as kindergarten, first grade, and second grade. In later grades the housing variable was not significant and therefore was left out of the model. The R^2 for the majority of the multiple regression equations was greater than 99.8% without the housing variable. From these results, it is evident that the housing variable did not add information in predicting the dependent variable.

To compare the results of all three techniques, an Absolute Mean Error Rate (AMER) was computed for each year. AMER is computed by taking the absolute value of the percent error in each grade and finding the mean of these percent errors. By taking the absolute value, the total deviation from the actual enrollments is computed rather than computing an error rate that is neutralized by positive and negative percent errors. Percent error for each grade level was calculated as follows:

$$\text{Percent Error} = \frac{\text{Predicted Enrollment - Actual Enrollment}}{\text{Actual Enrollment}} \times 100$$

In determining what is considered accurate for enrollment projections, Schellenberg & Stephens (1987) conducted a survey of 50 school planners and administrators who formulate enrollment projections. Approximately 67% of the planners aim for an overall error rate of 1% or less for a projection of one year into the future. Another 20% of school planners aim for an overall error rate of 0.5% or less for a projection of one year into the future. If enrollments are calculated for individual grades one year into the future, 67% attempt to keep error rates less than 4% per grade. For overall projections of two to three years into the future, an acceptable error rate was 3 to 4%, while for projections of four to five years into the future, an acceptable error rate was 4 to 5%.

Results

In Table 2 below, a comparison of the Absolute Mean Error Rates for the Cohort-Survival Ratio and the MRT for the years 1990 to 1996 in Districts A is presented.

Insert Table 2 about here

In District A for predicting enrollments from 1990 to 1996, the Modified Regression Technique using data from 1978 to 1989 had the best Absolute Mean Error Rate for a prediction of seven years into the future of 8.7%. Since CSR is the most commonly used technique presently, these results clearly demonstrate the superiority of using the MRT. The Cohort-Survival Ratio Method based on a two-year average ratio performed much poorer than the CSR based on a four year average but neither technique performed as well as the MRT in predicting four to seven years in the future. In the short-term, the CSR based on a four-year average ratio had the best AMER of all techniques when predicting one year into the future of 2.8%. In subsequent predicting years, the CSR based on a 4-year average sustained higher error rates than the MRT. As shown in Figure 1 below, both of the methods presented appear to have a larger AMER when predicting far into the future.

Insert Figure 1 about here

In District B for predicting enrollments for 1990 to 1996, the MRT using data from 1978 to 1989 had the best AMER for prediction of one to seven years into the future as shown in Table 3 below. The MRT using data from 1978 to 1989 had the best prediction of one year into of the future with an AMER of 2.0%.

Insert Table 3 about here

While the CSR four-year average had lower Absolute Mean Error Rates than the CSR using the two-year average, it did not provide more accurate results than either of the Modified Regression Technique models as shown in Figure 2 below.

Insert Figure 2 about here

In the long term prediction, the AMER for the MRT of 3.8% in 1996 is better than the CSR four-year average AMER of 5.3%. However, compared to District A, the difference in error rates between the MRT and the CSR using the four-year average was much lower. This is probably due to the computed Cohort-Survival Ratios being closer to the actual ratios from birth to kindergarten. All methods had increasing AMER as the number of years predicted into the future increased.

In comparing the AMER for the techniques employed for Districts A and B for predicting enrollments for 1993 to 1996 only, similar results were found but are not reported here. The MRT had lower Absolute Mean Error Rates than the Cohort-Survival Ratio Method for both districts and for both sets of data used. The results of these analyses are fully reported in Grip (1998).

In summary, there are substantial improvements in the Absolute Mean Error Rate when using the Modified Regression Technique when compared with the Cohort-Survival Ratio. The CSR based on a four-year average performed uniformly better than the CSR based on a two-year average. The MRT performed better than the CSR four-year average in District A in predicting from 1991 to 1996. Only in 1990 did the CSR four-year average outperform MRT. In District B, MRT performed better than the CSR four-year average in predicting from 1991 to 1996. In 1990, the CSR and MRT had the same Absolute Mean Error Rate. For the long-term predictions of seven years into the future, CSR seems to perform well in the first year of prediction, but performs poorly for enrollment predictions further into the future.

Discussion

In the first documented use of the Modified Regression Technique, this model outperformed the Cohort-Survival Ratio method for two suburban school districts in central New Jersey. The Modified Regression Technique provided lower Absolute Mean Error Rates than the Cohort-Survival Ratio Method for enrollment predictions from 1990 to 1996 or from 1993 to 1996. These findings are consistent with those made by Webster (1970) who used a single predictor variable, the number of current students in a particular grade, in a simple regression equation to calculate the number of students. As a comparison, Webster then performed enrollment projections using the Cohort-Survival Ratio for the same time period. In Webster's study, the simple regression model provided more accurate enrollment projections than the Cohort-Survival Ratio for a period of five to eight years into the future. The Modified Regression Technique has added two additional variables, the number of housing units and a migration ratio, to predict school enrollments with a greater accuracy.

When predicting enrollments for 1990 to 1996 using MRT, 12 (1978-1989) and 17 (1973-1989) years of data were used. Consistently, the twelve years of data produced the most accurate results. Similarly, when predicting enrollments for 1993 to 1996 using MRT, 15 (1978-1992) and 20 (1973-1992) years of data were used. The smaller data set, fifteen years, produced the most accurate results. From these results, a school planner should not necessarily pick the greatest number of years available, but instead use data that closely resembles the prevailing trends.

In general, the error rates were larger for District A using either method. For most accurate results, both the Cohort-Survival Ratio and the Modified Regression Technique work best under conditions of uniform growth or decline. From 1990-1996, District A had a rate of growth of nearly 54% while District B grew 25%. Since District A did not experience uniform growth, the error rates were larger for this district. Another reason District A had higher error rates is that the projected birth to kindergarten cohort ratios were much higher than the actual number of students progressing from birth to kindergarten. Since

both methods depend on the accuracy of this ratio, the accuracy of the enrollment projections was effected. However, it appears that, based on its lower Absolute Mean Error Rates, MRT is a more robust model than CSR for making predictions when the data trend is changing rapidly.

Since MRT was developed, the researchers have acted as consultants for various school districts across the nation. MRT has been used for districts much larger than those used in the study and the projections have been very close to what school administrators expected. Because MRT predicts up to seven years into the future, the effectiveness of the model will not be known for years to come when the actual student counts are compared with the projections. Since MRT has been applied to school districts that are different than those used in the study, it appears that MRT is a robust tool for performing enrollment projections.

Limitations of Using the Modified Regression Technique

One limitation in using the Modified Regression Technique is predicting the values of the independent variables such as the number of housing units in the district, the number of births, and the migration ratio. These values need to be predicted before the number of students in each grade can be calculated. In Districts A and B, the existing trend of birth rates and housing growth continued for the years which these variables were predicted. If, for example, interest rates had risen sharply during this period causing fewer houses to be built, the estimation for the housing variable would be inaccurate. The accuracy of the Modified Regression Technique is related to the accuracy in predicting the independent variables. Long-term predictions are especially volatile since many social and economic changes could occur during the period of prediction and alter the predicted enrollments significantly. Since CSR is dependent on the number of students progressing from one grade to the next remaining constant for the entire prediction period, this method is also very sensitive to any changes that occur during the prediction period.

REFERENCES

Ascher, W. (1978). Forecasting: An appraisal for policy-makers and planners. Baltimore: Johns Hopkins University Press.

Bernhardt, V. L., Pullum, T. W., Graham, S. N., (1983). Seattle's small-area approach to forecasting enrollments at the school level. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Quebec, Canada. (ERIC Document Reproduction Service No. ED 239 378)

Caffarella, E. P. (1983). Determining the accuracy of enrollment projection for school districts. CEFP Journal, 21(4), 14-16.

Caffarella, E. P. (1987). Effects of the baby boomette on school enrollment. School Business Affairs, 53(7), 26-28, 41-43.

Castaldi, B. (1989). Educational facilities, planning, remodeling and management. New York: Allyn and Bacon.

Chan, T. C. (1993). How many will choose? School choice and student enrollment planning. School Business Affairs, 59(7), 29-31.

Cline, D. H. (1985). Calculating the financial impact of population growth on education. Paper presented at the Annual Meeting of the Northern Rocky Mountain Educational Research Association, Jackson Hole, WY. (ERIC Document Reproduction Service No. ED 310 512)

Colombo, L. J., Dekker, D. R., & Petronis, J. (1990). Counting Heads. The American School Board Journal, 177(8), 20-21, 30.

Dekel, G. P., (1994). Projecting school enrollment: a solution to a methodological problem. School Business Affairs, 60(4), 32-34.

Glass, T. E., & Fulmer, C. L. (1991). Enrollment projections- factors and methods. CEFPI's Educational Facility Planner, 29(6), 25-30.

Grip, R. S. (1998) Prediction of public school enrollments using the Modified Regression Technique (Doc[illegible]al Dissertation, Rutgers University, 1998).

Harris, D. E., Burrage, P. M., & Smith, W. J. (1986). Local insights keep projections on the money. The Executive Educator, 8(11) 20-21.

Lee, S., & Shim, J. (1990). Micro management science: microcomputer applications of management science (2nd Ed.). Boston: Allyn and Bacon.

MacConnell, J. D. (1957). Planning for School Buildings. Englewood Cliffs, New Jersey: Prentice-Hall.

Murdock, S. H. & Ellis, D. R. (1991). Applied Demography. Boulder, Colorado: Westview Press.

Pfitzner, C. B. (1987). A short-term forecasting procedure for institutional enrollments. Community/Junior College Quarterly, 11(3), 141-152.

Schellenberg, S. J., & Stephens, C. E. (1987). Enrollment projection: variations on a theme. Paper presented at the Annual Meeting of the American Educational Research Association, Washington D.C., (ERIC Document Reproduction Service No. ED 283 879)

Shaw, R. C. (1984). Enrollment forecasting: What methods work best? NASSP Bulletin, 68(468), 52-58.

Webster, W. J. (1970). The Cohort-survival ratio method in the projection of school attendance. The Journal of Experimental Education, 39(1), 89-96.

Webster, W. J. (1971). An empirical comparison of selected enrollment projection formulas. The Journal of Educational Research, 64(8), 375-381.

Weldon W., Hurwitz, E., & Menacker J. (1989). Enrollment projections: techniques and financial implications. School Business Affairs, 55(10), 28-31.

Table 1

Ethnic Composition of District A and B

Race	District A[a]	District B[b]
White	67.62	81.80
Black	10.41	3.13
Hispanic	4.95	4.79
American Indian/Alaskan Native	0.03	0.02
Asian/ Pacific Islander	16.99	10.26
Total	100.00	100.00

Note.

[a] Composition is expressed as a percentage and is based on data from Superintendent's files from October 1997.

[b] Composition is expressed as a percentage and is based on data from Superintendent's files from October 1996.

Table 2

Comparison of Absolute Mean Error Rates in District A for 1990-1996 using the Cohort-Survival Ratio and Modified Regression Technique

	Method			
Year	CSR 2-Year Average	CSR 4-Year Average	MRT[a]	MRT[b]
1990	4.5	**2.8**	3.0	3.2
1991	7.6	4.9	4.4	**4.1**
1992	9.9	6.4	5.1	**5.0**
1993	12.9	8.2	**6.7**	**6.7**
1994	16.8	10.4	8.5	**8.1**
1995	19.3	11.3	9.3	**8.9**
1996	20.0	11.3	9.0	**8.7**

Note. Bolded values represent the lowest AMER for the year analyzed of all the techniques.
[a] Modified Regression Technique based on data from 1973 to 1989
[b] Modified Regression Technique based on data from 1978 to 1989

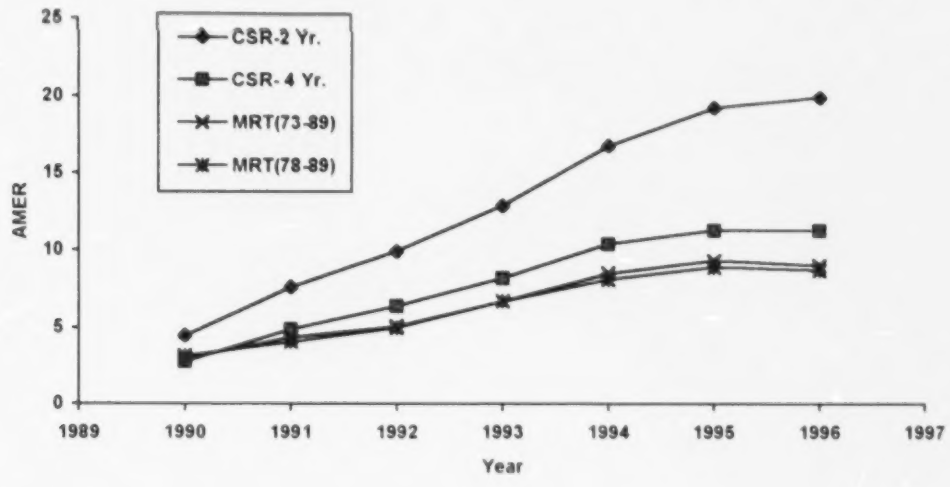

Figure 1 Absolute Mean Error Rates (AMER) for the predicted years 1990 to 1996 in District A using Cohort-Survival Ratios based on a two and four year average ratio and the Modified Regression Technique based on data from 1973 to 1989 and 1978 to 1989

Table 3

Comparison of Absolute Mean Error Rates in District B for 1990-1996 using the Cohort-Survival Ratio and Modified Regression Technique

	Method			
Year	CSR 2-Year Average	CSR 4-Year Average	MRT[a]	MRT[b]
1990	2.8	2.5	2.3	**2.0**
1991	3.6	3.2	3.6	**2.7**
1992	3.8	4.1	3.9	**2.8**
1993	3.7	3.8	4.5	**3.5**
1994	5.2	4.8	5.6	**4.1**
1995	5.8	5.2	6.2	**4.2**
1996	5.6	5.3	5.7	**3.8**

Note Bolded values represent the lowest AMER for the year analyzed of all the techniques

[a] Modified Regression Technique based on data from 1973 to 1989

[b] Modified Regression Technique based on data from 1978 to 1989

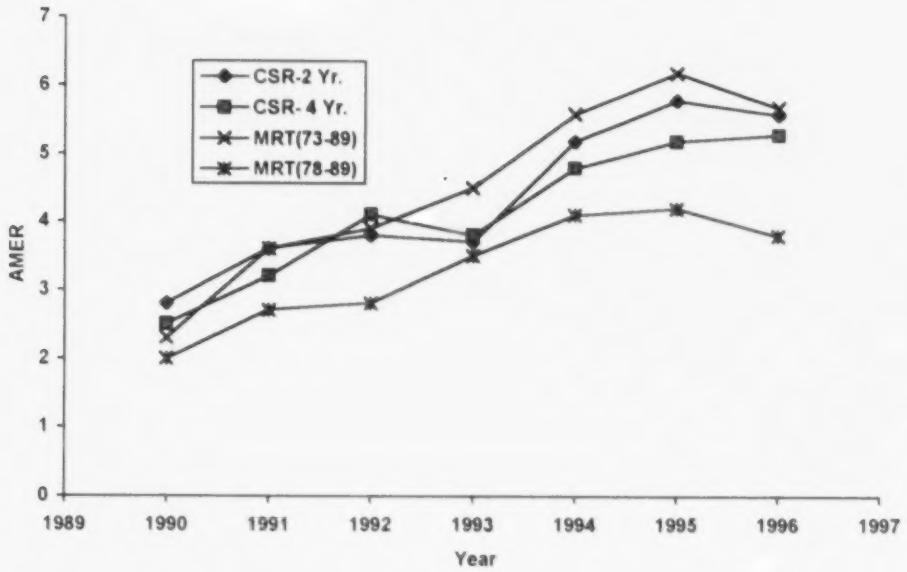

Figure 2. Absolute Mean Error Rates (AMER) for the predicted years 1990 to 1996 in District B using Cohort-Survival Ratios based on a two and four year average ratio and the Modified Regression Technique based on data from 1973 to 1989 and 1978 to 1989